DOI: 10.13137/2039-8646/37280

L'insegnamento della Statistica in ambito multidisciplinare - Parte seconda: Statistica e incertezza[*]

LUCIO TORELLI
Dipartimento Universitario Clinico di
Scienze Mediche, Chirurgiche e dalla Salute
Università di Trieste
torelli@units.it

ABSTRACT

Statistics can be an important tool to present in class, not only as a certain number of formulas to learn and apply, but as a set of elements that lead to reasoning in front of a problem. In "Teaching Statistics in a Multidisciplinary Area - Part One"[1], I presented some basic elements of descriptive Statistics, highlighting some critical issues that can emerge from statistics done in a rushed manner or not contextualized and therefore incorrect. In this second part, I will present some topics related to Probability. Statistics must consider uncertainty, which is present in data collection, in the choice of a sample, and so on. The Calculus of Probabilities is a necessary step for those who want to do statistics, as well as for being able to read critically what we see in data processing. Topics presented in class and that may seem too theoretical or useless, take shape in a different way in the face of statistical analyses that ask to be able to get information from the data in order to be able to make decisions. In this second part I will present biomedical applications that are of interest not only to those in the field. In fact, I will talk about Probability applied to diagnostic tests and risk measures. These topics, arising from the possibility of working as a mathematician in a multidisciplinary field, can be presented in class even as for independent lessons, to give some practical examples. Sharing questions and designing experiments together to reach some objectives, is an important team game, where everyone brings their specific contribution to the personal growth and of the entire team. I think that this way of working together can also be an important lesson for school students.

PAROLE CHIAVE

STATISTICA / STATISTICS; ELABORAZIONI STATISTICHE / STATISTIC; DIDATTICA INTERDISCIPLINARE / INTERDISCIPLINARY TEACHING; INCERTEZZA / UNCERTAINTY; TEST DIAGNOSTICI; DIAGNOSTIC TESTS; MISURE DI INCERTEZZA / UNCERTAINTY MEASURES.

---

[*] *Title: Teaching Statistics in a Multidisciplinary Context - Part Two: Statistics and Uncertainty.*

[1] Cfr. TORELLI 2024, pp. 7-24.

1. Introduzione

Da diversi anni insegno Matematica e Statistica medica e svolgo ricerche in collaborazione con colleghe e colleghi medici e sanitari. Sono stati e sono anni importanti, di conoscenza reciproca, umana e professionale. Le colleghe e i colleghi, le studentesse e gli studenti con cui lavoro, spesso apparentemente lontani dalle mie materie, le hanno scoperte in maniera nuova: non solo aridi strumenti di calcolo, ma discipline che possono aiutare a ragionare anche quando si affrontano problemi biomedici, in un percorso da fare insieme fin dalla progettazione di uno studio.

Viviamo, anche in ambito professionale, di tante statistiche, ma non sempre si conoscono in maniera appropriata gli strumenti da utilizzare. Molte statistiche sono fatte in maniera incompleta o affrettata, senza un progetto o senza una chiara metodologia con cui lavorare. Non basta avere un certo numero di dati per poter fare una statistica, così come non basta fare la media dei dati raccolti per poter descrivere in maniera corretta una certa situazione.

Nel precedente contributo che ho dedicato a questi argomenti[2], ho fatto alcuni esempi di statistiche che, pur corrette dal punto di vista algebrico, non sono contestualizzate nella situazione che si intende presentare. In quel testo ho fatto anche una breve presentazione di elementi di base di *Statistica descrittiva* e mi riservavo una seconda parte per presentare alcuni argomenti legati al *Calcolo delle Probabilità*[3].

È mio desiderio affrontare questi argomenti cercando di arrivare a tutti, a persone con preparazioni e conoscenze diverse e non solo strettamente legate alla Matematica. Mi sembra una sfida importante, in quanto c'è sempre più bisogno di *saper dialogare* tra discipline diverse, mantenendo le proprie specificità. È una realtà che può diventare una proposta didattica importante, che può aprire la mente alle studentesse e agli studenti, anche in vista di quelle che saranno le scelte lavorative di un domani. Una caratteristica essenziale della Statistica è la sua capacità di cercare di far fronte

[2] Cfr. Torelli 2024, pp. 7-24.
[3] Nella terza parte affronterò alcuni elementi di base di *Statistica inferenziale.*

all'*incertezza.* Mi sembra allora importante dare qui qualche cenno su questo tema, anche per offrire spunto per l'insegnamento nelle nostre Scuole. Non farò quindi un'introduzione classica al Calcolo delle Probabilità, che rimando a molti testi validi[4], ma presenterò delle tematiche partendo da alcune applicazioni.

Si dice che la Statistica certa sia un ossimoro; la Statistica, infatti, deve tenere conto dell'incertezza presente in vari aspetti del processo di raccolta dei dati: nella scelta di chi raccoglie i dati, di quando e come questi vengono raccolti, della loro numerosità e così via. Fare ad esempio una stima puntuale, dare cioè un unico valore, può risultare comodo ma non esprime l'incertezza della stima stessa.

Sapere che un amico arriverà a un appuntamento alle 19, non fornisce un'informazione completa e, infatti, di solito si aggiunge l'avverbio "circa": la *stima puntuale* viene così sostituita da una *stima intervallare*. Nello specifico, l'amico potrebbe precisare che arriverà tra le 19 e le 19:10, tra le 19 e le 20 o altro ancora e, di fronte a un intervallo più grande, prenderò una decisione diversa per come organizzarmi nell'attesa.

Ci piacciono le classifiche, i ranking e spesso li chiediamo erroneamente alla Statistica: sapere in un sondaggio pre-elettorale che un candidato è al 34% delle preferenze mentre un altro è al 29% ci fa pensare che tra i due ci sia una distanza di 5 punti percentuali, come se stessimo leggendo la classifica di un campionato sportivo, ma in realtà non è così, in quanto si tratta di stime calcolate a partire da dati campionari. Infatti, se supponiamo che il primo candidato abbia preferenze con una stima tra il 30% e il 38%[5] e il secondo tra il 26% e il 32%, notiamo che l'intersezione tra i due intervalli non è vuota e pertanto il sondaggio dice che la situazione non è chiara e che c'è anche qualche probabilità che il secondo candidato sia in vantaggio rispetto al primo.

In questo testo darò alcuni suggerimenti partendo da due semplici applicazioni del calcolo delle probabilità, apparentemente di interesse solo in ambito biomedico, ma in realtà di uso quotidiano da parte di tutti noi: parlerò dei *test diagnostici* e delle

[4] Cfr. Rossi 2022, Dall'Aglio 2003, Baldi 2011.
[5] In gergo giornalistico si parla di "forchetta di valori".

misure di rischio. È interessante far scoprire come argomenti apparentemente esclusivi di certe discipline, utilizzino in realtà elementi importanti di matematica.

2. LE TAVOLE DI CONTINGENZA

Prima di parlare dei test diagnostici è bene riprendere in mano le cosiddette *tavole di contingenza*, strumenti molto utili nel descrivere due variabili. Mi è successo più volte di presentare in aula un problema di questo tipo: la tavola di contingenza di Figura 1 (in questo caso a due righe e a due colonne, ma potrebbero presentarsi per altre variabili più righe e/o più colonne) riporta i dati di pazienti sopra e sotto i 75 anni di età, 200 e 300 rispettivamente, che hanno avuto o meno una recidiva, 100 recidive contro 400, dopo un certo trattamento. Ci chiediamo quale sia la probabilità di recidiva all'interno di tutto il gruppo di 500 pazienti e, a seguire, la probabilità di recidiva solo tra i pazienti over 75.

	Recidiva Sì	Recidiva No	
più di 75	50	150	200
meno di 75	50	250	300
	100	400	500

Figura 1. Esempio di tavola di contingenza.

Di solito le studentesse e gli studenti rispondono in maniera corretta, anche se non hanno mai studiato tavole di questo tipo e se non hanno grandi conoscenze di probabilità. Ci chiediamo, in aggiunta, se l'informazione di considerare solo persone del gruppo sopra i 75 anni modifichi in maniera significativa la probabilità iniziale richiesta. È un modo semplice per introdurre il *concetto di probabilità condizionata* e per iniziare a parlare di *indipendenza* tra eventi e tra variabili.

Nel caso qui riportato la probabilità di recidiva all'interno di tutto il gruppo è uguale a 1/5 (100 recidive su un totale di 500 pazienti), mentre se consideriamo solo gli ultra settantacinquenni è 1/4 (50 ultra settantacinquenni con recidiva, su un totale di 200).

Considerare solo persone più anziane ha modificato, in base ai dati analizzati, la valutazione della probabilità di recidiva: i dati raccolti, come si è visto, dicono che la probabilità di recidiva nel gruppo è del 20% (1/5) e sale al 25% (1/4) se ci si restringe alla fascia più anziana. In maniera analoga notiamo che la proporzione tra pazienti con recidiva e pazienti senza recidiva è 100 su 400 nel campione totale, ma è diversa nelle altre righe della tavola, dove leggiamo 50 su 250 e 50 su 150.
Questa breve premessa sulle tavole di contingenza ci permette di introdurre, da un punto di vista matematico, i *test diagnostici.*

3. Probabilità e misura dell'incertezza nei test diagnostici

Pochi anni fa, in tempo di COVID, si è parlato molto di kit diagnostici: per sapere se eravamo stati contagiati si faceva infatti un test sierologico, sostituito successivamente da quello molecolare, per passare infine anche a test salivari, meno costosi e più rapidi nel fornire un risultato. Si parlava e si parla di accuratezza dei kit diagnostici: ora cerchiamo di approfondire il significato di questo termine, che forse abbiamo usato in maniera un po' inconsapevole e non precisa, e cerchiamo di capire come mai questi argomenti sono delle applicazioni del calcolo delle probabilità.

	M+	M-	
T+	veri positivi	falsi positivi	positivi
T-	falsi negativi	veri negativi	negativi
	malati	sani	totale

Figura 2. Descrizione di una tavola di contingenza per un test diagnostico.

	M+	M-	
T+	20	30	50
T-	5	145	150
	25	175	200

Figura 3. Esempio di dati raccolti per un test diagnostico.

Le tavole di contingenza di Figure 2 e 3 sono un modo semplice per la rappresentazione di un *test diagnostico*, in cui abbiamo indicato, per un certo test T, con T+ i risultati positivi, con T- quelli negativi, con M+ la presenza della malattia che si vuole testare e con M- l'assenza della stessa malattia.
Le colonne e le righe dei totali riportano il numero dei malati, 25 in questo esempio, e dei sani, 175, delle persone positive al test, 50, e di quelle negative, 150. Le celle interne alla tavola riportano nella diagonale il numero dei veri positivi, 20 e dei veri negativi, 145, cioè il numero di persone per cui il test ha dato risposte corrette; nell'antidiagonale troviamo invece il numero di falsi positivi, 30, persone positive al test ma in realtà sane, e quello dei falsi negativi, 5, persone negative al test ma in realtà malate.
I dati raccolti dicono che l'informazione sull'esito del test modifica la probabilità di essere malati: risulta infatti pari a 20 su 50, 40%, la probabilità di essere malati sapendo di essere risultati positivi, contro una probabilità del 12,5%, 25 su 200, di essere malati nell'intero gruppo.
Questa *probabilità condizionata*, che viene matematicamente indicata con la notazione $P(M+|T+)$[6] viene chiamata *valore predittivo positivo* Vp+, in quanto predice la malattia, noto l'esito del test. In maniera analoga si definisce il *valore predittivo negativo* Vp-, cioè la probabilità di essere sani, sapendo di aver avuto esito negativo dal test[7]. I Valori predittivi vengono tenuti solitamente separati; succede infatti che per alcune malattie si preferisca avere dei falsi positivi piuttosto che dei falsi negativi. In altre situazioni potrebbe essere invece preferibile il contrario.
Esistono altre due definizioni probabilistiche che riguardano i test e che servono a valutare la bontà di un kit diagnostico: la *sensibilità* e la *specificità*. Da un punto di vista matematico la cosa è molto semplice: nelle definizioni dei valori predittivi si scambia la variabile T con M, operazione non commutativa che genera due nuove definizioni ossia: $S_e = P(T+|M+)$, detta *sensibilità*, che valuta la probabilità di risultare positivi sapendo

[6] E si legge "Probabilità di essere malati sapendo di essere positivi al test" (in inglese si dice, ancor meglio, *within positive).*
[7] Probabilità condizionata pari a 96,7%, data da $\frac{145}{150}$.

che si è malati, e S_p = P(T-|M-), detta *specificità*, ovvero la probabilità di risultare negativi essendo sani. Può sembrare un semplice passaggio teorico matematico, inutile dal punto di vista assistenziale[8], ma in realtà questi ultimi concetti sono utili proprio per valutare la bontà di un kit.

Supponiamo di voler valutare un nuovo test diagnostico e supponiamo di avere un gruppo di persone che sappiamo, per altra via, essere malate e un gruppo di persone sane cui somministriamo il test. La sensibilità misura, sui dati del campione scelto, l'efficienza del test nell'individuare come positivi i malati. In maniera analoga accade per la specificità, se si somministra il test a un certo numero di persone sane. I valori stimati della sensibilità e della specificità forniscono pertanto l'accuratezza del kit diagnostico.

Nel caso della tabella in Figura 3, risultano le seguenti stime: sensibilità= 80,0%[9], calcolata dividendo il numero dei veri positivi per quello dei malati, specificità= 82,9%[10], ottenuta dividendo il numero dei veri negativi per quello dei sani. Un test con sensibilità pari al 100% non produrrà falsi negativi in quanto, dalla definizione, una persona malata risulterà certamente positiva. Un test con specificità del 100% non permetterà la comparsa di falsi positivi.

Riassumendo

Vp+ = P(M+|T+) = (veri positivi)/(positivi)

Vp- = P(M-|T-) = (veri negativi)/(negativi)

valutano le probabilità di malattia, dato l'esito del test

Se = P(T+|M+) = (veri positivi)/(malati)

Sp = P(T-|M-) = (veri negativi)/(sani)

valutano le caratteristiche di un test

---

[8] Se sappiamo che una persona è malata viene predisposta una cura e di solito non interessa fare un altro test.

[9] Dato da $\frac{20}{25}$.

[10] Dato da $\frac{145}{175}$.

È importante, infine, tenere conto del fatto che il risultato del test, positivo o negativo, dipende dalla scelta di un valore soglia, detto *cut off*.[11] A un valore del cut off più basso corrispondono più positivi - e quindi meno falsi negativi - e in modo analogo a un valore di cut off più alto corrispondono più negativi - e quindi meno falsi positivi. Per individuare un valore di cut off, può essere utilizzata la cosiddetta curva R.O.C., un grafico che mostra i valori di *sensibilità* e *specificità* al variare dei possibili cut off[12].

La Figura 4 mostra una curva R.O.C. con sei diversi valori soglia e con i corrispondenti valori di specificità e di sensibilità (in realtà si è soliti mettere in ascissa (1 – specificità) e in ordinata la sensibilità). Il test ideale, con sensibilità e specificità pari al 100%, si trova nel vertice in alto a sinistra nel grafico[13]; verrà pertanto scelto uno dei nodi della spezzata più vicini a tale vertice, a seconda che si voglia un valore più alto di sensibilità o di specificità.

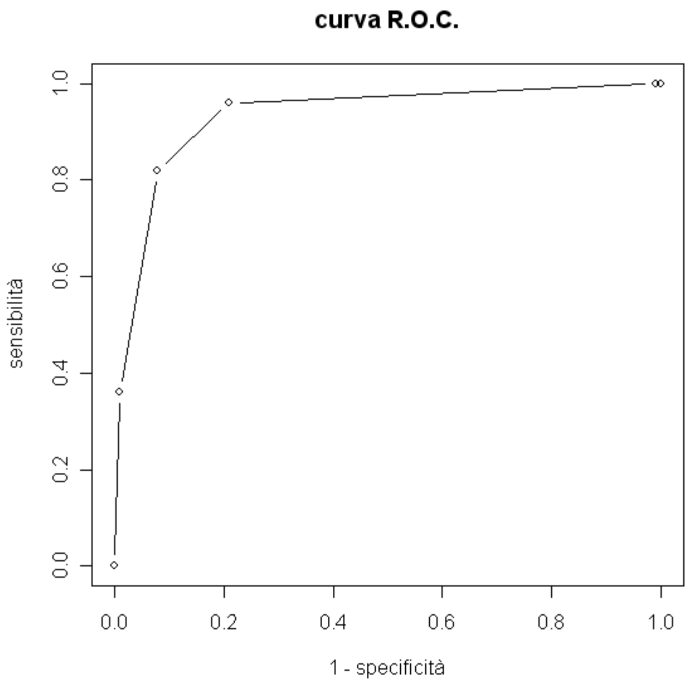

Figura 4. Esempio di curva R.O.C.

[11] Quando andiamo, ad esempio, a fare degli esami del sangue, il risultato delle analisi fornisce il valore cercato e anche l'informazione se tale valore è inferiore o superiore a un valore limite prefissato.

[12] Interessante la storia della curva R.O.C. (*Receiver Operating Characteristic*), che si racconta sia nata dopo l'attacco di Pearl Harbour nella seconda guerra mondiale: i radar americani non furono in grado di riconoscere gli aerei giapponesi, considerati come dei falsi negativi. Da lì alcuni ingegneri cercarono degli strumenti che permettessero di individuare un cut off migliore.

[13] Punto di coordinate (0;1), con (1-specificità)=0, ovvero specificità=1 e sensibilità=1.

Propongo ora un esercizio utile per capire meglio l'interesse di tali concetti. Questo nasce da un episodio realmente accaduto: una mia conoscente era risultata positiva a un certo test di screening e, preoccupata, aveva chiesto se potevo avere qualche spiegazione da un amico medico che si occupava di questo programma di screening. La risposta del medico fu incoraggiante in quanto disse, in base alla sua esperienza, che molto probabilmente si sarebbe trattato di un falso positivo. Si sarebbe fatto comunque un accertamento, ma ci disse che la cosa molto probabilmente si sarebbe risolta bene. E così fu.

In me rimaneva comunque il dubbio dell'utilità di tale test e così provai a fare una *simulazione*. Avevo le informazioni sul test, con sensibilità circa del 60% e specificità del 90%. Il collega medico aggiunse che in quelle situazioni di screening, la *prevalenza* della malattia[14] cercata era circa dell'1%. Con questi dati dovevo ricavare il valore predittivo positivo, la probabilità che la mia conoscente fosse realmente malata, sapendo che era risultata positiva.

Come affermato in precedenza, la sensibilità può essere calcolata, sui dati di un opportuno campione, come il rapporto tra il numero di veri positivi e il totale di malati e la specificità come rapporto tra il numero di veri negativi e il totale dei sani. In questo modo si possono trovare le percentuali di veri positivi e veri negativi e di conseguenza ricavare una stima dei valori predittivi[15].

da Se = (veri positivi)/(malati) $\Rightarrow$ (veri positivi) = Se $*$ (malati)
da Sp = (veri negativi)/(sani) $\Rightarrow$ (veri negativi) = Sp $*$ (sani)

con Se = 60%, Sp = 90%, M+ = 1%

La percentuale di veri positivi si può ricavare come prodotto della sensibilità e della percentuale di malati, ottenendo un valore stimato pari a 0,6% (=60%$*$1%) e in maniera

[14] Cioè la percentuale di individui malati in una popolazione, in un determinato momento.

[15] Da sensibilità=(veri positivi)/malati, ricaviamo che (veri positivi)=Sensibilità$*$malati, etc.

analoga per i veri negativi, dal prodotto della specificità e della percentuale di sani, si ottiene 89,1% (=90%*99%). I falsi positivi e negativi verranno quindi calcolati per differenza: a questo punto, come si vede nella Figura 5, è possibile stimare, in base ai dati in ingresso, il valore predittivo positivo cercato, dato dal rapporto tra i veri positivi e i test positivi, ottenendo Vp+=5,7%. In maniera analoga si calcola il Valore predittivo negativo Vp-=99,6%.
Cosa rispondere a questo punto alla mia conoscente? Come leggere questi due risultati ottenuti con alcuni semplici passaggi algebrici?

	M+	M-	
T+	0,6%	9,9%	10,5%
T-	0,4%	89,1%	89,5%
	1%	99%	100%

Figura 5. Simulazione dei risultati di un test diagnostico con Se=60%, Sp=90%, M+=1%.

Il valore predittivo positivo calcolato conferma quanto aveva affermato il medico, in base alla sua esperienza: una persona positiva a tale test ha una probabilità molto bassa, inferiore al 6% di essere malata[16]. Il valore predittivo negativo è invece molto alto: con una probabilità vicina al 100% una persona negativa al test non è malata[17].
Esercizi di questo tipo possono aiutare a simulare gli esiti di nuovi kit diagnostici prima di decidere su un loro eventuale acquisto: se una Azienda di prodotti biomedicali propone dei nuovi prodotti con determinate caratteristiche, possiamo verificarne da subito la bontà, per un loro utilizzo, senza aspettare di provarli su un numero opportuno di pazienti.

4. Probabilità e misure del rischio

Non è raro sentire parlare di un fattore come un *rischio* rispetto al verificarsi di un

[16] Si è passati da una probabilità di malattia dell'1% a una di quasi il 6%, che rimane comunque bassa.
[17] Entra comunque sempre in gioco la professionalità del medico che usa questi risultati come uno strumento, insieme ad altri, per valutare la salute del paziente.

certo *evento*: seguire una dieta non corretta espone a rischi per alcune malattie, fattori genetici e di familiarità aumentano il rischio di particolari patologie, l'esposizione alla polvere di amianto ha aumentato in maniera pesante e drammatica il rischio di asbestosi e di mesotelioma e, al contrario, una dieta bilanciata e sana - così come pure l'abitudine di camminare quotidianamente - protegge da alcune malattie, l'allattamento al seno per almeno sei mesi protegge da alcuni tumori, e così via.

Sono notizie che talvolta possono generare una cattiva informazione, se non vengono presentate e capite in maniera corretta[18]. Sapere che due eventi sono tra loro *dipendenti* può essere un primo elemento di valutazione, come nell'esempio sopra riportato sulla recidiva rispetto alle fasce d'età, ma ora si tratta di fare di più, di cercare di capire un'eventuale correlazione di causa-effetto e di misurarne il rischio.

Numero di sigarette	Bronchite SI	Bronchite NO	Totali
50 o più di 50	5	2	7
tra 20 e 49	15	11	26
tra 5 e 19	10	12	22
tra 1 e 4	20	25	45
0	50	250	300
	100	300	400

Figura 6. Dati relativi a fumo e bronchite.

Supponiamo di avere raccolto i dati riportati nella tabella di Figura 6. Il *Rischio Relativo* (RR) viene definito come il rapporto tra le probabilità condizionata di essere malato (in questo caso di avere la bronchite) tra gli esposti al fumo rispetto a quella di avere la bronchite tra chi non è esposto. Pertanto:[19]

$$RR=P(M+|\text{esposto})/P(M+|\text{non esposto}) = (50/100)/(50/300) = 3$$

[18] Recentemente un articolo della California University, poi ripreso in maniera intelligente dal podcast *Ci vuole una scienza*, si chiedeva quanto alto fosse il rischio di fare troppi esami diagnostici e riportava la domanda «quante TAC sono troppe TAC?» (cfr. MENIETTI E., MAUTINO in Siti web e SMITH-BINDMAN *et al.* 2025).

[19] I fumatori con bronchite sono infatti 50, dati dalla somma di 5, 15, 10 e 20, per un totale di fumatori pari a 100, somma di 7, 26, 22, 45. I non fumatori con bronchite sono invece 50 su un totale di 300.

Chi fuma ha quindi un rischio triplo, in base ai dati del campione in oggetto, di avere la bronchite cronica[20]. Rimane però il dubbio di come sia stato definito "essere fumatori", nei cosiddetti materiali e metodi dello studio[21]. Come per i test diagnostici, i risultati dipendono dalle scelte metodologiche fatte all'inizio dello studio.

	Bronchite SI	Bronchite NO	
5 o più di 5	30	25	55
meno di 5	70	275	345
	100	300	400

Figura 7. Fumo e bronchite con *cut off* pari a 5 sigarette.

Supponiamo che i dati ora siano stati catalogati in due classi, come in Figura 7 e non nelle cinque classi della Figura 6: rispetto a tale nuova scelta metodologica in cui pensiamo a un fumatore come qualcuno che fuma 5 o più di 5 sigarette al giorno, il rischio relativo assume il valore RR=2,7: sono infatti 30 su 55 gli ammalati nel primo gruppo contro 70 su 345. Facendo invece un'altra scelta metodologica in cui vengono confrontati i grandi fumatori (50 o più di 50 sigarette al giorno) contro chi non fuma[22], come da Figura 8, Il Rischio Relativo diventa RR=4,3.

	Bronchite SI	Bronchite NO	
50 o più di 50	5	2	7
0	50	250	300
	55	252	307

Figura 8. Fumo e bronchite con 50 o più di 50 sigarette e non fumatori.

Possiamo pertanto generare risultati diversi a seconda delle scelte metodologiche che

[20] I dati dicono che per la probabilità di avere la bronchite per l'intero gruppo è pari a 100/400 = 25%, ed è del 50% = 50/100 nei fumatori, del 16,7% = 50/300 nei non fumatori.
[21] Nei materiali e metodi andrà dichiarato come devono essere contati gli ex fumatori o chi è soggetto a fumo passivo.
[22] È una scelta che spesso viene fatta quanto si vogliono trovare risultati che facciano notizia, che colpiscano l'opinione pubblica, senza evidenziare le scelte metodologiche con cui si è deciso di confrontare i due estremi di una certa situazione.

assumiamo nei nostri studi. La Statistica deve essere uno strumento che aiuta a ragionare sull'elaborazione dei dati e sulla loro contestualizzazione. Questo è uno dei motivi per cui può succedere che alcuni studi vengano presentati con risultati apparentemente estremi, ma che vanno invece considerati all'interno delle scelte metodologiche fatte.

In letteratura si trova spesso un altro concetto di misura di rischio, l'*Odds Ratio* (OR), che qui non approfondiremo e tipico solitamente di studi retrospettivi[23], dove con Odds di un evento si intende il rapporto tra la sua probabilità e la probabilità che l'evento non accada[24].

Sapere che un certo fattore è un rischio, ad esempio, doppio per l'insorgenza di una malattia, può essere un elemento importante o no, a seconda di quanto questo valore va a moltiplicare: se il rischio di una malattia è molto basso, sarà bassa anche la probabilità di chi è esposto al fattore, ma se il rischio è più alto, situazione tipica ad esempio di pazienti geriatrici o fragili, allora il moltiplicatore 2 porterà a una probabilità importante di malattia.

Da questi esempi viene sempre più in evidenza che la Statistica non è solo un insieme di formule che portano a un risultato, ma uno strumento utile se inserito e contestualizzato nel problema che si sta affrontando.

5. Alcuni pensieri per concludere questa seconda parte

Abbiamo visto come elementi di Calcolo della Probabilità, che spesso in classe vengono affrontati con considerazioni didatticamente interessanti, ma apparentemente inutili, quali lanci di monete o di dadi, possono invece portare ad applicazioni importanti e di impatto sociale. Ad esempio, fino a poco tempo fa si era infatti discusso molto sui diversi tipi di tampone da utilizzare per il COVID; o, ancora, sapere che un certo fattore aumenta il rischio di una malattia può portare a cercare di modificare stili di vita e a prestare maggiore attenzione a modalità di prevenzione e di cura.

---

[23] Il RR è solitamente usato negli studi prospettici.

[24] Lanciando un dado possiamo dire la probabilità che esca 2 è 1 su 6, ma possiamo anche dire che è 1 contro 5, modalità usata ad esempio nei sistemi di scommesse.

Sono esempi che di solito interessano le studentesse e gli studenti, perché fanno constatare l'utilità della Matematica e della Statistica e facilitano una maggiore comprensione degli argomenti che si studiano in classe. Sono, inoltre, esempi e argomenti che aiutano gli insegnanti a dare agli studenti una visione più aperta, multidisciplinare della realtà, senza perdere la specificità di ogni sapere. Queste caratteristiche, tra l'altro, sono sempre più richieste anche nel mondo del lavoro, mondo che gli studenti si troveranno poi ad affrontare.

Nella terza parte di questo lavoro, si farà infine una breve introduzione a elementi di base di *Statistica inferenziale*, anch'essa presentata partendo da esempi e applicazioni, cercando di essere comprensibili anche per coloro che non hanno approfondite conoscenze in questo campo.

BIBLIOGRAFIA

BALDI P.
2011, *Calcolo delle Probabilità*, Milano, McGraw Hill.

DALL'AGLIO G.
2003, *Calcolo delle Probabilità*, Bologna, Zanichelli.

ROSSI L.
2022, «Probabilità, che problema», *QuaderniCIRD,* 24 (2022), pp. 10-28.

SMITH-BINDMAN R., CHU P. W., AZMAN FIRDAUS H. *et al.*
2025, «Projected Lifetime Cancer Risks From Current Computed Tomography Imaging», *JAMA Intern* Med., 185(6), pp. 710–719, scaricabile dall'indirizzo:
<https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2832778>.

TORELLI L.
2024, «L'insegnamento della Statistica in ambito multidisciplinare. Parte prima», *QuaderniCIRD*, 29 (2024), pp. 7-24.

SITI WEB

MENIETTI E., MAUTINO B.
Quante TAC sono troppe TAC?,
<https://www.ilpost.it/podcasts/ci-vuole-una-scienza/quante-tac-sono-troppe-tac/>, sito consultato il 9.6.2025.